

Effect of Ethnicity on the Speaking Fundamental Frequency Values of Indians vis-a-vis Caucasians, Afro-Americans and Mongoloids: A Forensic Phonetic Perspective

Dr. Swapna Akoju¹, Dr. Grace Suneetha Didla²

¹Lecturer, Tribal Welfare Degree College, Bhadrachalam, Telangana, India

²Asst. Prof., Department of Phonetics, The English & Foreign Languages University, Hyderabad, Telangana, India

Abstract— It is established in the forensic literature that the ethnic background of an individual has a robust bearing on one's speaking fundamental frequency (SFF). The aim of this study is to record the SFF values of Indian adult speakers and compare the SFF of Indian speakers with those of existing data on other ethnic groups like Caucasians, Afro-Americans and Mongoloids in order to examine the effect of ethnicity on SFF. In addition, it is also to observe if language and mode of speech has any bearing on the SFF values. The study included 20 Indian speakers whose ages ranged from 21 to 40 years. A read passage and spontaneous speech in two languages (Telugu and English) constituted the text. PRAAT software was used to extract SFF values and a comparison was made between the obtained SFF values of Indian speakers and those drawn from other studies on three ethnic groups: Caucasians, Afro-Americans and Mongoloids. Results indicate that the Indian speakers (both male and female) exhibited a marginally higher pitch (on an average) compared to those of the other three ethnicities.

Keywords— Afro-Americans, Mongoloids, SFF, PRAAT.

I. INTRODUCTION

Documenting any aspect of speech is pivotal to understanding human language, culture and behavior. Speech signals not just the linguistic message but also the indexical properties of the speaker. Identification of the indexical properties of speech may be referred to in the forensic literature as 'speaker profiling'. Speaker profiling entails identifying the age, gender, region, ethnicity or socio-economic background of the speaker based on the speech. Of the many phonetic factors that lend to speaker profiling, one such robust factor which is of relevance to this study is the identification of the ethnicity of an individual based on their speaking fundamental frequency (a long-term average of pitch). Since globalization has opened up doors to multi-culturalism and multi-ethnicity, identifying the ethnicity of an individual is of paramount significance in the field of forensic phonetics. Another application of speaker profiling in recent times is the Linguistic analysis of the determination of origin (LADO) in cases of refugees seeking asylum, where indexical

properties play a significant role in determining their claimed ethnicity.

1.1. Forensic Speaker Identification

Forensic phonetics is a fledgling discipline in the domain of forensic linguistics which deals with the identification of criminals based on the phonetics aspects of speech such as the segmental and suprasegmental features. It deals with several aspects such as speaker identification, voice line-ups, tape-authentication, speaker profiling etc.

In speaker identification "an utterance from an unknown speaker has to be attributed, or not, to one of a population of known speakers for whom reference samples are available" [1]. Identifying people based on their speech has gained significance in the recent past.

There exist many factors which can positively affect the speaker identification process such as: large and good quality speech samples, familiarity with the speaker, listener's talent, phonetic training and structured and validated analysis [2]. However, there are quite a few features that can mar the speaker identification process,

viz., multiple speakers, voice disguise, stress, text independent samples, differing health states, alcohol & drugs, differing dialects, sound-alikes and noise [3 & 4].

1.2. Speech Correlates of Forensic Speaker

Identification: There are several segmental and suprasegmental features of speech that assist in speaker identification.

1.3. Segmental Features

These include the analysis of vowel and consonant sounds. A thorough perceptual and acoustic analysis of consonants and vowels can aid in speaker identification. Vowels are produced with a continuous airflow through the vocal tract which makes them predominantly voiced. On the other hand, consonants are produced by causing obstruction to the air flow and include both voiced and voiceless sounds. The feature of voicing is the result of the vibration of the vocal folds. Vowels play a significant role in the process of forensic speaker identification, because their acoustical properties are relatively strong and easy to quantify or measure. Rose [5] points out that “vowels are prominent not only because they last longer, have greater duration than consonants but also because of their relatively well-defined acoustic structure. The acoustic properties of vowels show the imprint of the vocal tract through which they have been produced.” Apart from this, vowel quality also plays a major role in identifying the accent of a speaker. Accent is an important feature in identifying a person’s dialect. On establishing the dialect of the speaker, it is easy to profile him/her to a specific social group and geographical area.

1.4. Suprasegmental Features

Suprasegmental features include stress, tone, intonation, pitch, etc. While stress is the relative prominence of syllables within a word; tone is the use of different pitch shapes to signal word identity; whereas, Intonation is the use of pitch contour on longer utterances [5]. All these suprasegmental features are primarily monitored by pitch. Since Pitch, the perceptual correlate of Fundamental Frequency (F0), is a quasi-permanent feature of an individual’s speech and is also easy to extract from connected speech, it plays a major role in the forensic speaker identification.

Fundamental frequency (F0) is the number of vocal fold vibrations per second. Fundamental frequency predominantly depends upon the mass, length and tension of the vocal folds. F0 can be effectively extracted through sustained vowels. Hence, most of the research on F0 has been carried out on sustained vowels. Nolan [1] observes

that measures associated with F0 have shown to be among the more successful in speaker recognition. In addition, Rose [5] points out that F0 is robust and can be extracted very easily from the recordings of a poor quality. However, he cautioned against the indiscriminate use of F0 through several factors which affect within-speaker variation in F0. Based on Braun’s [6] views, Rose [5] categorized the factors as physical (race, age, smoking and intoxication), psychological (emotional state) and technical factors (sample size and tape speed). Besides these factors, other situational factors like background noise also have a bearing on the changes in F0 patterns.

Speaking Fundamental Frequency (SFF), which is a long-term average of fundamental frequency, is identified as a better parameter compared to the ‘Fundamental Frequency’ in the speaker identification process. Nolan [1] expounded that the average SFF plays a major role in the speaker identification process. The same view has been echoed by Hollien [7] who says that the SFF is one of the primary features of speech which aids in the process of speaker recognition. Here the focus is laid on the mean pitch values of spoken form. It is common knowledge that no person speaks with a monotonous pitch. The pitch patterns vary depending upon the vibrations of the vocal folds. During speech, the vocal folds vibrate in different patterns and produce different ranges of frequencies, sometimes lower and sometimes higher. Hence, in conversational or connected speech, SFF plays a major role than F0 as it indicates the central tendency of one’s pitch in speech.

Apart from segmental and suprasegmental features, aspects like Voice Quality, Tempo, Vocal Intensity and General Speech also assist the investigator in the process of speaker identification. ‘Voice Quality’ refers to the uniqueness of an individual’s voice. Just as two types of musical instruments sound differently though the same note is played with the same intensity, likewise, two speakers may resemble each other (as in the case of twins), yet their voice quality may be different. In addition, the ‘Tempo’ of speech also aids in the speaker identification process. A person can fairly be identified by how slow or fast and how smooth or choppy his/her speech is. On the other hand, ‘Vocal Intensity’ is contributed by the sub-glottal air pressure that is exerted when a person speaks. Intensity varies depending upon the distance between the speaker and the microphone; therefore, it is difficult to assess the speaker by his or her vocal intensity alone. In addition, several features of ‘General Speech’ like, dialect, idiosyncratic pronunciation and language patterns, the unusual use of linguistic stress and speech impediments also play a major role in the process of speaker identification.

II. BACKGROUND LITERATURE

That 'Pitch' is a good indicator of speaker's identity [5], and ethnicity brings a change in the pitch levels [7] has been established in the forensic literature. In the light of this, quite a substantial work has been carried out on the pitch values of different ethnic groups in the world such as: Caucasians [8 & 9], Afro-Americans [10] and Mongoloids [11], to name a few. A brief summary of these studies is presented below.

2.1. Effect of Ethnicity on the SFF values of Children With an aim to observe the differences in SFF among children, Awan & Mueller [12] examined the SFF values of 105 children (3-6 years) belonging to three different ethnic groups. The subjects included 35 White speakers (15 boys & 20 girls), 35 African-American speakers (18 boys and 17 girls), and 35 Hispanic speakers (16 boys & 19 girls). The following table reflects the SFF values obtained from the study.

Table.1: SFF Values of Children across three different ethnicities

S.No.	Ethnicity	SFF values	
		Boys	Girls
1	Whites	240.07 Hz	243.35 Hz
2	African-American	241.31 Hz	231.48 Hz
3	Hispanic	248.99 Hz	248.04 Hz

The study indicated that there was a moderate effect of ethnicity on the SFF values of these speakers.

2.2. Effect of Ethnicity on the SFF values of Adolescents

There have been quite a few studies which compared the SFF values of adolescent speakers of one ethnic group with another. In this regard, Hollien-Malcik team [13] experimented on 18 Southern Negro (SN) boys of 3 different age groups [10, 14 and 18 years]. The results revealed that these boys exhibited SFF values of 223 Hz, 163 Hz and 124 Hz respectively. Subsequently, the SFF values of these boys were compared with those of Northern White (NW) boys (as reported by Curry [14], which is represented below.

Table 2: SFF values of Northern White boys and Southern Negro boys

S.No	Age	Ethnicity	SFF
1	10 years	NW	270 Hz
		SN	223 Hz
2	14 years	NW	242 Hz
		SN	163 Hz
3	18 years	NW	137 Hz
		SN	124 Hz

As is evident, the SFF values of Northern White boys are higher than those of Southern Negroes across all age groups. The study reported that 'the blacks experienced voice change earlier than the whites do'. Yet another study [15] was carried out to observe the influence of climatic conditions on the SFF values. In this study, data on 491 boys residing in four different countries (150- Swedish, 180- Dutch or Polish and 161- Spanish) was collected to test the hypothesis that climate change might be a factor that influences the adolescent voice change (AVC). This study proved that AVC seemed to occur earlier in Swedish boys, who were from a cold climate than from the Dutch boys, who were from a temperate region. The AVC occurred much later in the Spanish boys compared to these two countries.

2.3. Effect of Ethnicity on the SFF values of other age groups

Natour and Wingate [16] carried out a study on the SFF values of 300 Jordanian Arabic speakers (both adults and children) observed values are as follows: 137.45 Hz (male speakers), 230.84 Hz (female speakers) and 278.04 Hz (children). On comparison of the obtained values with those of other ethnicities such as Caucasians and African-Americans, results indicate that the SFF values of male and female speakers of Jordanian Arabic were similar. However, the children of Jordanian Arabic have exhibited higher SFF values than those of Caucasian children.

Yet another interesting study on Japanese men and women was carried out by Nishio & Niimi [11]. The study included 374 Japanese speakers (divided into 3 groups: young adults, middle aged and old aged) who were asked to render "The North Wind and the Sun" passage in Japanese. On the extraction of SFF values from their speech samples, results indicate that the mean SFF values of male speakers were 121.83 Hz (young adults), 120.95 Hz (middle aged) and 127.82 Hz (old aged) and those of

female speakers were 224.58 Hz (young adults), 196.31 Hz (middle aged) and 178.92 Hz of the aged women.

The study further indicated that the older women exhibited a noteworthy decrease in their SFF values. Given below is a graphical representation of SFF values of Japanese male and female speakers across different age groups.

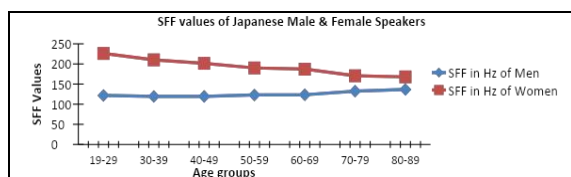


Fig.1: SFF values of Japanese Male & Female speakers across different age groups

It is intriguing to note that with the advancement in age, a moderate increase in the SFF values of men was observed. However, a substantial decrease was observed in the SFF values of female speakers.

III. PRESENT STUDY

While there have been scores of papers capturing the SFF values of several ethnicities of the world, there is acute dearth of research on the SFF values of Indian speakers. Although moderate attempts have been carried out by pathologists in restricted domains, a forensic phonetic perspective was rarely attempted.

3.1. Aim

The aim of this study is to record the SFF values of Indian adult speakers and compare the SFF of Indian speakers with those of existing data on other ethnic groups like Caucasians, Afro-Americans and Mongoloids in order to examine the effect of ethnicity on SFF. In addition, it is also to observe if language and mode of speech has any bearing on the SFF values.

3.2. Methodology

3.2.1. Choice of Speakers

20 Indian bilingual speakers (categorized into 2 age groups: 21-30 and 31-40) were chosen for the study. The speakers' age was between 21 and 40 years. All the subjects were bilinguals whose mother tongue was Telugu (a popular South-Indian Language) and whose second language was English.

3.2.2. Choice of Text

An ideal and practical way to examine the speech was to make the speakers read and talk spontaneously. Therefore, two modes of speech were used during the process of recording. Every speaker was asked to read out a passage

in English and Telugu titled: 'The North Wind and the Sun' and 'Kaki-Kadava Passage' respectively. Since reading is completely different from speaking, the subjects were asked to talk spontaneously for a minute in each of the languages, (English and Telugu) on one of the following topics: 'The person they like most', 'the movie they like most' or 'the game they like most'. The two modes of speech served a binary purpose.

While the chosen text in English (of 196 words) took approximately one and a half minutes for rendering, the text in Telugu (of 103 words) lasted about one minute. The said passages were chosen for recording since they were phonetically balanced.

3.3. Analysis

PRAAT software was used in extracting the long-term pitch patterns from each of the recorded speech samples. The obtained Pitch values were compared with the Pitch values drawn from other studies on three ethnic groups: Caucasians, Afro-Americans and Mongoloids.

3.4. Findings

3.4.1. Normative data on the SFF Values of Male Speakers

Since the primary aim was to gather normative data on the SFF of Indian bilingual speakers, speech samples were collected in two languages (English and Telugu) and in two modes of speech (Read and spontaneous). Given below is a table which summarizes the normative data on the SFF values of 21-40 years male speakers. The SFF values shown in the table are the averages of their read passage and spontaneous speech (English & Telugu).

Table 3: SFF values of Male Speakers across different age groups

Age Group	Read speech		Spontaneous speech		Average Values
	English	Telugu	English	Telugu	
21-30	135.8	135	135.8	133.6	135.05
31-40	126.6	125.4	125.8	123.4	125.3

It may be observed that within each age group neither the language nor the mode of speech had any effect on the SFF values of the speakers. However, it may be noted that across the two age groups, the 21-30 year age group exhibited a marginal increase in SFF values compared to the 31-40 year age group.

3.4.2. Normative data on the SFF Values of Female Speakers

The table summarizing the normative data on the SFF values of Indian bilingual female speakers of age group 21-40 years has been represented below. As mentioned earlier, the SFF values shown in the table are the average values of their read passage and spontaneous speech in two languages (English and Telugu).

Table 4: SFF values of Female Speakers across different age groups

Age Group	Read speech		Spontaneous speech		Average Values
	English	Telugu	English	Telugu	
21-30	218.6	218.8	212.2	213.2	215.7
31-40	220	215.2	218.6	211.6	216.3

The data on the female speakers reveals that age, language and mode of speech had no impact on the SFF values of the speakers.

3.4.3. Effect of Ethnicity on SFF Values of Adult speakers

Since this study also aimed at comparing the SFF values of Indian speakers with those of the existing literature on several other ethnicities in the world, given below is a comprehensive graph summarizing the SFF values of 21-40 year old speakers drawn from several studies varying over different periods of time.

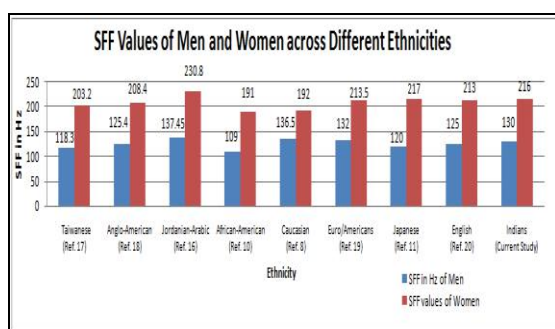


Fig 2: SFF values of adult Men and Women across different ethnicities.

The graph represents nine different studies dealing with several ethnic groups. Among men, the Jordanian-Arabs recorded the highest value of 137 Hz, while the African-Americans exhibited lowest SFF value of 109 Hz. It is evident from Study-9 (which is the current study) that the SFF values of Indian bilingual speakers differed marginally, when compared to speakers from other ethnic backgrounds.

It is interesting to note that the same pattern was observed even among the female speakers. While the highest SFF

values were those of Jordanian-Arabic speakers, the lowest values were of the African-Americans.

IV. CONCLUSION

The following is a summary of the conclusions drawn from the study.

- 4.1. The normative data on the SFF values of bilingual Indian men is recorded as 135.05 Hz (21-30 years group) and 125.3 (31-40 years group). It further reveals that across both age groups, neither the language nor the mode of speech had any affect on the SFF values of the speakers. However, it may be noted that 21-30 year age-group exhibited marginally higher SFF values compared to the 31-40 year age-group.
- 4.2. The normative data on the SFF values of the bilingual Indian women is recorded as 215.7 Hz (21-30 years group) and 216.3Hz (31-40 years group). It further reveals that age, language and mode of speech had absolutely no bearing on their SFF values.
- 4.3. On comparing the SFF values of male speakers from several ethnicities, it may be noted that the Jordanian-Arabic speakers recorded the highest value (137 Hz) while the African-American speakers exhibited the lowest value (109 Hz).
- 4.4. On comparing the SFF values of female speakers from several ethnicities, the same trend was observed as that of males. While the Jordanian-Arabic speakers recorded the highest value (230.8 Hz) while the African-American speakers exhibited the lowest value (191 Hz).

REFERENCES

- [1] Nolan, F. (1983). The Phonetic Bases of Speaker Recognition, Cambridge: Cambridge University Press.
- [2] Hollien, H. & Majewski, W. (1997). As cited in Hollien, H. (1990) The Acoustics of Crime, New York: Plenum Press.
- [3] Broeders, A. P. A. (1995). The role of automatic speaker recognition techniques in forensic investigations. Proc. Intl. Congress Phonetic Sciences 3: 154-61.
- [4] Campbell, J., Shen, W., Campbell, W., Schwartz, R., Bonastre, J.F. & Matrouf, D. (2009). Forensic Speaker Recognition. IEEE Signal Processing Mag., March: 95-103.
- [5] Rose, P. (2002). Forensic Speaker Identification, Taylor & Francis.
- [6] Braun, A. (1995). Fundamental Frequency- How Speaker Specific Is It. Braun and Koster (eds): 9-23.
- [7] Hollien, H. (1990). The Acoustics of Crime, New York: Plenum Press.

- [8] Murry, T., Brown, W., & Morris, R. (1995). Patterns of fundamental frequency for three types of voice samples. *Journal of Voice*, 9:282-289.
- [9] Hollien, H. & Shipp, T. (1972). Speaking fundamental frequency and chronologic age in males. *Journal of Speech and Hearing Research*, 15:155-157.
- [10] Hudson, A. I. & Holbrook A. (1982). Fundamental Frequency Characteristics of Young Black Adults. *JSHR*, 25:25-28.
- [11] Nishio, M. & Niimi, S. (2008). Changes in the Speaking Fundamental Frequency Characteristics with Aging. *Folia PhoniatrLogop*, 60:120-127.
- [12] Awan, S. & Mueller, P. (1996). Speaking Fundamental Frequency Characteristics of White, African American and Hispanic Kindergartners. *JSHR*, 39:573-577.
- [13] Hollien, H. & Malcik, E. (1962). Adolescent Voice Change in Southern Negro Males. *Speech Monogr*: 53-58, as cited in Spienza CM, (1997).
- [14] Curry, et al. (1940). The pitch characteristics of the adolescent male voice. *Speech Monogr.*, 7:48–62.
- [15] Hollien, H. & Hollien, P. A. (1972). A cross-cultural study of adolescent voice change in European males. *Proceedings of 7th International Congress Phonetic Sciences*, The Hague, The Netherlands: Mouton and Co: 332–357.
- [16] Natour, Y. S. & Wingate, J. M. (2009). Fundamental Frequency Characteristics of Jordanian Arabic Speakers. *Journal of Voice*, 23:5.
- [17] Wang, C. C. & Huang H. T. (2004). Voice Acoustic Analysis of Normal Taiwanese Adults. *Journal of Chinese Medical Association*, 67(4):179-184, as cited in *Journal of Voice* (2009) 23:560-566.
- [18] Hollien, H. & Jackson, B. (1973). Normative data on the speaking fundamental frequency characteristics of young adult males. *Journal of Phonetics*, 1(2):117-120.
- [19] Fitch, J. L. & Halbrook, A. (1970). Modal Vocal Fundamental Frequency of Young Adults. *Archives of Otolaryngology*, 92:379-382.
- [20] Graddol, D. (1986). Discourse Specific Pitch Behaviour, in *Intonation in Discourse*, edited by Catherine Johns-Lewis, 221-237.